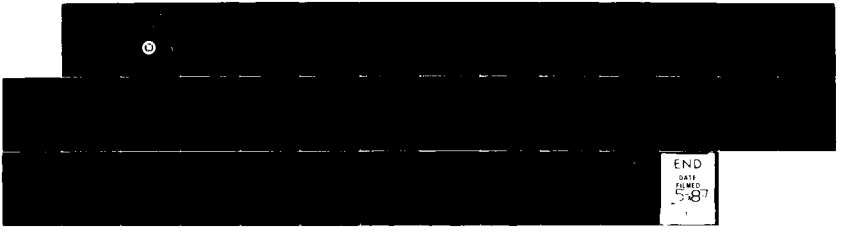
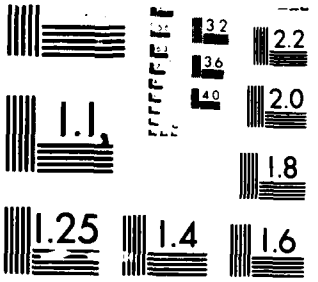
AD-A178 540 ESTIMATION OF THE PARAMETERS OF A MODIFIED COMPOUND POISSON DISTRIBUTION..(U) RENSSELAER POLYTECHNIC INST TROY NY SCHOOL OF MANAGEMENT J L BRYANT ET AL. 1986 1/1

UNCLASSIFIED WP-37-86-P21 ARO-18072.17-MA F/G 12/1 NL



END
DATE FILMED
5-87



Rensselaer Polytechnic Institute
School of Management

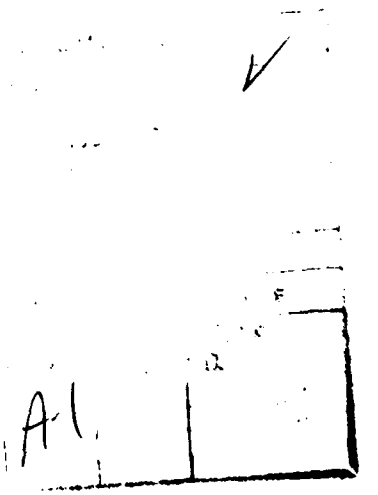
Working Paper No. 37-86-P21

ESTIMATION OF THE PARAMETERS OF A MODIFIED COMPOUND POISSON DISTRIBUTION

by

J. L. Bryant
and
A. S. Paulson

School of Management
Rensselaer Polytechnic Institute
Troy, NY 12180-3590
(518)266-6586



Estimation of the Parameters of a Modified Compound Poisson Distribution

by

J.L. Bryant

and

A.S. Paulson*

* Research sponsored in part by U.S. Army Research Office Contract DAAG29-81-K-0110.

Summary

This paper addresses the problem of estimation of the parameters of the Poisson sum of Gaussian random variables imbedded in a background of Gaussian noise when only realizations of the sum are observable. Cumulant matching, maximum likelihood, and an empirically orthogonalized characteristic function procedures are considered. The characteristic function and the maximum likelihood procedures produce similar results in a simulation study. However, the characteristic function procedure is computationally superior. Conditions under which all procedures are incapable of parameter estimation are discussed.

Key words: Characteristic function, empirical characteristic function, compound Poisson, maximum likelihood, normal variables.

1. Introduction

Let the random variable Z_0 have a normal distribution with zero mean and variance σ_1^2; denote by $Z_1, Z_2, \ldots$, a sequence of variates which are identically normally distributed with means μ and variances σ_2^2. The Z_j, $j=0,1,2,\ldots$, are taken to be independent of one another and also of the discrete random variable N, which has a Poisson distribution with parameter λ. The problem of concern in this paper will be the estimation of the parameters $\underset{\sim}{\theta} = (\mu,\lambda,\sigma_1^2,\sigma_2^2)'$ when only realizations of the sum

$$X = Z_0 + \sum_{k=1}^{N} Z_k \tag{1.1}$$

are observable.

The topic will be motivated in this section by a brief account of a certain security price model proposed by Press (1967;1968) which leads to the estimation problem at hand. Models of this type occur frequently in communications engineering and can also be categorized as being of the cumulative damage or asset flow type, so that it is likely to be of interest in a wide variety of possible applications.

The fundamental assumptions of the price fluctuation model advocated by Press may be summarized by supposing that the net increase or decrease in value of a security over a given time interval may be represented as a random sum of independent price changes superimposed on an independent process of background noise. Each price change is triggered by the arrival of some "information event," which occurs from time to time in accordance with a Poisson process N(t) having parameter λ. The logged price of the security (which should be adjusted to compensate for stock

splits, divided payments, and so forth) at some time t can then be characterized by the equation

$$P(t) = P_0 + \sum_{k=1}^{N(t)} Z_k + Y(t), \quad t \geq 0, \tag{1.2}$$

where P_0 is the initial log-price at the base time t=0, the Z_k are independent random variables representing price changes due to the occurrence of information events, and Y(t) is the background noise process, Y(0)=0. Press takes the Z_k to be normal with mean μ and variance σ_2^2 and supposes that Y(t) is a Wiener process with parameter σ_1^2, so that Y(t) has stationary and independent increments, and for any t>0 Y(t) has a normal distribution with zero mean and variance $\sigma_1^2 t$. The processes N(t) and Y(t) are assumed to be independent of one another and of the Z_k. P(t) represents the log-price of the security rather than the price itself primarily to account for the empirically justifiable belief that the variation of price change should be positively related to the magnitude of a security's value.

Security price data are typically compiled at regular time intervals, whose length may be taken to be one unit without loss of generality. Then, letting X_t, t=1,2,..., represent the change in log-price of the security in the interval (t-1,t], it follows by differencing equation (1.2) that

$$X_t = \sum_{k=N(t-1)+1}^{N(t)} Z_k + Z_{0,t} \qquad t=1,2,..., \tag{1.3}$$

where $Z_{0,t} = Y(t) - Y(t-1)$, and so is normally distributed with mean zero and variance σ_1^2. Note also that the number of terms in the summation of equation (1.3) is N(t) - N(t-1), which is therefore Poisson distributed with parameter $\lambda \Delta t = \lambda$. Thus X_t is distributed exactly as the random variable

X of equation (1.1). Furthermore, the log-price changes X_t and $X_{t'}$ are independent for $t \neq t'$ since the processes Y(t) and N(t) have independent increments and the random variables $Z_1, Z_2, \ldots$, have been assumed independent. If a realization of the process P(t), $0 \leq t \leq n$ is available, then the problem of estimating its parameters $\underset{\sim}{\theta} = (\mu, \lambda, \sigma_1^2, \sigma_2^2)'$ may therefore be reduced to the problem of estimating the parameters of the distribution associated with the variate X of (1.1), based on the random sample $X_1, X_2, \ldots, X_n$.

Empirical investigations of Fama (1965) indicate that the distribution of log-price changes should possess thicker tails and be more peaked about some measure of central tendency than would be permitted by a Gaussian distribution. Press' compound events model, represented by equations (1.2) and (1.3), can be shown to possess these properties (Press, 1968). It is also in general skewed, a property which some empirical evidence suggests may be appropriate (Fielitz and Smith, 1972; Leitch and Paulson, 1975).

2. Estimation by Cumulant Matching

By a conditioning argument, it is easy to show that the distribution function associated with the random variable X of (1.1) is

$$F(x, \underset{\sim}{\theta}) = e^{-\lambda} \sum_{q=0}^{\infty} \frac{\lambda^q}{q!} \Phi\left(\frac{x-q}{\left(\sigma_1^2+q\sigma_2^2\right)^{\frac{1}{2}}}\right), \tag{2.1}$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

is the distribution function of a standard normal deviate; the corresponding

density is

$$f(x;\underset{\sim}{\theta}) = \frac{e^{-\lambda}}{\sqrt{2\pi}} \sum_{q=0}^{\infty} \frac{\lambda^q}{q!} \frac{\exp\left\{-\frac{(x-q\mu)^2}{2(\sigma_1^2+q\sigma_2^2)}\right\}}{\left(\sigma_1^2 + q\sigma_2^2\right)^{\frac{1}{2}}} \quad ; \tag{2.2}$$

the corresponding characteristic function is

$$\phi(u;\underset{\sim}{\theta}) = \exp\{-\tfrac{1}{2}\, u^2\sigma_1^2 + \lambda(e^{i\mu u - \frac{1}{2}u^2\sigma_2^2} - 1)\}. \tag{2.3}$$

The cumulants of the distribution may be found by developing the cumulant generating function $\log \phi(u;\underset{\sim}{\theta})$ in powers of u. We shall require the first four of these:

$$\begin{aligned} \kappa_1 &= \lambda\mu \\ \kappa_2 &= \sigma_1^2 + \lambda(\mu^2 + \sigma_2^2) \\ \kappa_3 &= \lambda\mu(\mu^2 + 3\sigma_2^2) \\ \kappa_4 &= \lambda(\mu^4 + 6\mu^2\sigma_2^2 + 3\sigma_2^4) \end{aligned} \tag{2.4}$$

Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from (2.1). Generally there will be no need to explicitly consider the underlying process P(t) of equation (1.2) which may have generated the sample.

Since the density $f(x;\underset{\sim}{\theta})$ of (2.2) has no simple closed form, it can be seen that the method of maximum likelihood may not provide a computationally attractive solution to the problem of estimation of $(\mu,\lambda,\sigma_1^2,\sigma_2^2)$. For this reason and because of the simplicity of (2.4), Press (1967;1968) has suggested a cumulant matching procedure.

Let $\hat{\kappa}_j$, j=1,2,3,4, represent the first four cumulants of the sample $\{X_1,X_2,\ldots,X_n\}$. These are related to the sample mean $\bar{X}$ and the central moments $m_\nu = \frac{1}{n}\sum_{j=1}^{n}(X_j-\bar{X})^\nu$ by

$$\hat{\kappa}_1=\bar{X},\ \hat{\kappa}_2=m_2,\ \hat{\kappa}_3=m_3,\ \hat{\kappa}_4=m_4-3m_2^2. \tag{2.5}$$

Equating the sample cumulants to the respective population cumulants given in equation (2.4) yields a system of four equations in the four unknown parameters. After some reduction the system may be written as

$$\begin{aligned}
&\mu^4 - \left(\frac{2\hat{\kappa}_3}{\hat{\kappa}_1}\right)\mu^2 + \left(\frac{3}{2}\frac{\hat{\kappa}_4}{\hat{\kappa}_1}\right)\mu - \left(\frac{\hat{\kappa}_3^2}{2\hat{\kappa}_1^2}\right) = 0\\
&\lambda = \hat{\kappa}_1/\mu \qquad (2.6)\\
&\sigma_2^2 = \frac{\hat{\kappa}_3-\mu^2\hat{\kappa}_1}{3\hat{\kappa}_1}\\
&\sigma_1^2 = \hat{\kappa}_2 - \left(\frac{\hat{\kappa}_1}{\mu}\right)(\mu^2+\sigma_2^2).
\end{aligned}$$

Cumulant matching estimates $\tilde{\theta}_n = (\tilde{\mu}_n,\tilde{\lambda}_n,\tilde{\sigma}_{1n}^2,\tilde{\sigma}_{2n}^2)'$ may then be defined by requiring that they satisfy the system of equation (2.6). The quartic equation has, of course, four roots, real or complex; in every case to which this procedure has been applied, it has been found that exactly two of these are real, and are of opposite sign. The root to which $\tilde{\mu}_n$ should be equated is then that real root which causes the intensity parameter estimate $\tilde{\lambda}_n = \frac{\hat{\kappa}_1}{\tilde{\mu}_n}$ to be positive; that is $\hat{\kappa}_1$ and $\tilde{\mu}_n$ should be of similar sign.

The sample cumulants $\hat{\kappa}_j$ are, apart from $\hat{\kappa}_1$, not unbiased estimators of the corresponding population cumulants, although they are of course consistent. They have therefore been replaced in (2.6) by the

first four of Fisher's k-statistics, which are unbiased (Kendall and Stuart, I, 1969, p. 281).

The most attractive feature of cumulant matching is its ease of application; one has only to compute the first four cumulants (or k-statistics) of the sample and then solve a quartic equation to obtain estimates which have the desirable properties of consistency and asymptotic normality. Unfortunately, such estimators appear to possess rather low efficiencies. In an analysis of the price fluctuations of ten securities used in computing the Dow Jones Industrial average, Press (1967) found that the use of the cumulant matching method led to infeasible estimates (with either $\tilde{\sigma}_1^2$ or $\tilde{\sigma}_2^2$ negative) in every case. After the infeasible parameter estimate was set equal to zero, he furthermore found that the distribution function computed by substitution of the estimated parameter values into equation (2.1) gave a visibly poor fit when graphically compared to the empirical distribution function associated with the sample $X_1, X_2, \ldots, X_n$. These estimates were based on sample sizes ranging from 185 to 499, and Press concluded that much larger sample sizes are required to achieve reasonable estimates by cumulant matching. This is consistent with the simulation results to be subsequently presented which indicate that for samples of size 500, cumulant matching is totally inadequate, at least in those portions of the parameter space that were considered.

3. Characteristic Function Estimators

Because the characteristic function $\phi(u,\underset{\sim}{\theta})$ associated with the modified compound Poisson distribution has a reasonably simple form, it was believed that a method of parameter estimation based on characteristic functions might provide estimators for the parameter vector $\underset{\sim}{\theta} = (\mu,\lambda,\sigma_1^2,\sigma_2^2)'$ with reasonable efficiency and computational tractability. Estimates for the true parameter vector $\underset{\sim}{\theta}_0$ can be obtained by numerically determining the zeros of

$$S_{n\theta} = 2\ \mathrm{Re} \sum_{j=1}^{n} \int_{-\infty}^{\infty} \frac{\partial\phi(u;\underset{\sim}{\theta})}{\partial\theta} (\phi(u;\underset{\sim}{\theta}) - \exp iux_j)^* |\psi(\phi(u;\underset{\sim}{\theta}))|^2\ du = 0 \qquad (3.1a)$$

for $\theta = \mu,\lambda,\sigma_1^2,\sigma_2^2$ and some function $\psi(\cdot)$. Clearly $E(S_{n\theta}) = 0$ under mild regularity conditions and we can expect that the M-estimators derived from this system or its discrete counterpart,

$$S_{n\theta}^{+} = 2\ \mathrm{Re} \sum_{j=1}^{n} \sum_{q=1}^{p} \frac{\partial\phi(u_q;\underset{\sim}{\theta})}{\partial\theta} (\phi(u_q;\underset{\sim}{\theta}) - \exp iu_q x_j)^* |\psi(\phi(u_q;\underset{\sim}{\theta}))|^2 = 0 \qquad (3.1b)$$

for some p and $\theta = \lambda,\mu,\sigma_1^2,\sigma_2^2$, will be consistent and asymptotically normal (Thornton and Paulson, 1977). An appealing feature of (3.1) is that the weight function adapts itself to the data under the purview of the assumed model, in this case the modified compound Poisson distribution. Apart from the weighting, equations (3.1) are very similar in form to the normal equations of nonlinear least squares

$$\sum \frac{\partial \text{ expected}}{\partial\theta} (\text{expected} - \text{observed}) = 0,$$

for parameter θ. The observed terms are replaced by $\exp iux_j$ and the expected terms are replaced by $\phi(u;\underline{\theta})$. It thus makes sense to choose the weights $\psi(\phi(u_q;\underline{\theta}))$ in inverse proportion to the standard deviations of the $\{\phi(u_q;\underline{\theta}) - \exp iu_q x_j\}$. However, the residuals $\{\phi(u_q;\underline{\theta}) - \exp iu_q x_j\}$ are not uncorrelated. It will therefore often be advantageous to make use of this correlation and we shall do so presently.

It will be more convenient to work with the quantity

$$y(u;\underline{\theta}) = \text{Re } \phi(u;\underline{\theta}) + \text{Im } \phi(u;\underline{\theta}), \tag{3.2}$$

and its sample estimate

$$\hat{y}_n(u) = \text{Re } \hat{\phi}_n(u) + \text{Im } \hat{\phi}_n(u) \tag{3.3}$$

with

$$\hat{\phi}_n(u) = n^{-1} \sum_{j=1}^{n} \exp iux_j = n^{-1} \sum_{j=1}^{n} (\cos ux_j + i \sin ux_j), \tag{3.4}$$

instead of $\phi(u;\underline{\theta})$. Clearly $E(\hat{\phi}_n(u)) = \phi(u;\underline{\theta})$ and hence $E(\hat{y}_n(u)) = y(u)$ for all u. It is easy to show that the covariance kernel of the real process $\hat{y}_n(u)$ is given by

$$\begin{aligned} K(u,v) &= n \operatorname{cov}(\hat{y}_n(u),\hat{y}_n(v)) \\ &= \text{Re } \phi(u-v;\underline{\theta}) + \text{Im } \phi(u+v;\underline{\theta}) - [\text{Re } \phi(u;\underline{\theta}) + \text{Im } \phi(u;\underline{\theta})][\text{Re } \phi(v;\underline{\theta}) \\ &\qquad + \text{Im } \phi(v;\underline{\theta})] \end{aligned} \tag{3.5}$$

(See also Bryant, unbpublished Ph.D. dissertation, Rensselaer Polytechnic Institute, 1977.) The residuals $\hat{y}_n(u) - y(u;\underline{\theta})$ have covariance kernel $n^{-1} K(u,v)$. Define

$$\hat{\underline{y}}_n = (\hat{y}_n(u_1), \hat{y}_n(u_2), \ldots, \hat{y}_n(u_p))^{\sim} \tag{3.6}$$

and

$$\underline{y}(\underline{\theta}) = (y(u_1,\underline{\theta}), y(u_2,\underline{\theta}), \ldots y(u_p,\underline{\theta}))^{\sim} \tag{3.7}$$

where $u_1, u_2, \ldots, u_p$ have been chosen so that the matrix $\underset{\sim}{K}$ is positive definite. It is always possible to choose such $u_1, u_2, \ldots, u_p$ since the covariance function K(u,v) is positive definite (Feller 1968, Ch. XIX). The p×1 random vector $\underset{\sim}{z}(\underset{\sim}{\theta}) = \underset{\sim}{K}^{-\frac{1}{2}}(\hat{\underset{\sim}{y}}_n - \underset{\sim}{y}(\underset{\sim}{\theta}))$ has covariance matrix $\underset{\sim}{I}$. The estimation $\underset{\sim}{\theta}$ may be effected through consideration of the objective function

$$Q(\underset{\sim}{\theta}) = (\hat{\underset{\sim}{y}}_n - \underset{\sim}{y}(\underset{\sim}{\theta}))'\underset{\sim}{K}^{-1}(\hat{\underset{\sim}{y}}_n - \underset{\sim}{y}(\underset{\sim}{\theta})) =$$

$$\sum_{q=1}^{p} \sum_{q'=1}^{p} (\hat{y}_n(u_q) - y(u_q,\underset{\sim}{\theta}))(\hat{y}_n(u_{q'}) - y(u_{q'},\underset{\sim}{\theta}))k^{qq'} \qquad (3.8)$$

where the $k^{qq'}$, q, q'=1,2,...,p, are elements of the inverse matrix $\underset{\sim}{K}^{-1}$.

Since $(\hat{\underset{\sim}{y}}_n - \underset{\sim}{y}(\underset{\sim}{\theta}))$ is asymptotically p-dimensional Gaussian, $Q(\underset{\sim}{\theta})$ is asymptotically χ^2 on p degrees of freedom. Accordingly, estimation of the parameters $\mu, \lambda, \sigma_1^2, \sigma_2^2$ by way of $Q(\underset{\sim}{\theta})$ is approximately a χ^2 minimum procedure and can be expected to be quite efficient. In fact, this has been independently and recently shown by Feurverger and McDunnough (1981). The function $Q(\underset{\sim}{\theta})$ depends on the unknown value $\underset{\sim}{\theta}_0$ of the parameters vector through the matrix $\underset{\sim}{K}$. Estimation could still be effected by regarding $\underset{\sim}{K}$ as a function of the minimizing variable $\underset{\sim}{\theta}$ but such a procedure would require an inversion of the matrix at each iteration of the minimizing algorithm and so would lead to computational expense. An alternative to direct minimization of the χ^2 - like statistic is to proceed in stages via a modified χ^2 minimum procedure where the matrix $\underset{\sim}{K}$ is held constant during the differentiation stage and allowed to be variable thereafter. Instead, we use the fact that the vector $\hat{\underset{\sim}{y}}_n$ is the mean of the independent and identically distributed random vectors s_j, j=1,2,...,n, whose elements are

$$s_{jq} = \cos(u_q X_j) + \sin(u_q X_j), \qquad q=1,2,\ldots,p,$$

where $X_1, X_2, \ldots, X_n$ is the sample drawn from the population whose parameters are to be estimated. The matrix $\underset{\sim}{K}$ in (3.8) may be replaced by the sample covariance matrix $\hat{\underset{\sim}{K}}_n$ with the general element

$$\hat{k}_{qq';n} = \frac{1}{n-1} \sum_{j=1}^{n} \{s_{jq} - \hat{y}_n(u_q)\}\{s_{jq'} - \hat{y}_n(u_q)\}.$$

Thus the characteristic function estimates $\hat{\underset{\sim}{\theta}}_n$ may be generated by minimizing over $\underset{\sim}{\theta}$ the sum

$$O_n(\underset{\sim}{\theta}) = \sum_{q=1}^{p} \sum_{q'=1}^{p} (\hat{y}_n(u_q) - y(u_q, \underset{\sim}{\theta}))(\hat{y}_n(u_{q'}) - y(u_{q'}, \underset{\sim}{\theta}))\hat{k}_n^{qq'} \qquad (3.9)$$

where the $\hat{k}_n^{qq'}$ are elements of $\hat{\underset{\sim}{K}}_n^{-1}$. After such estimates are obtained, they may be refined by using equation (3.5), evaluated at the estimated parameter values, to re-approximate the covariance matrix $\underset{\sim}{K}$; then a second minimization step may be performed.

The algorithm outlined in the preceding paragraph has been applied to simulated data having the modified compound Poisson distribution through the use of a simplex minimization procedure (Jacoby, Kowalik and Pizzo, 1972, p. 79) applied to equation (3.9), where the variables λ, σ_1^2 and σ_2^2 were replaced by their logarithms to result in an unconstrained problem. Discussion of the performance of the estimation procedure will be deferred until after likelihood is discussed. A total of p=40 points u_q were used, twenty placed symmetrically on either side of the origin; the effect of their placement was not extensively studied, but did not appear to be too critical, so long as several points were always included near the origin. Since

$\phi(\frac{u}{\sqrt{\kappa_2}},\underset{\sim}{\theta})$, where κ_2 is the variance of the distribution given by (2.2), corresponds to random variables with unit scale, it seems reasonable to use the quanity $\frac{1}{\sqrt{\hat{\kappa}_2}}$, where $\hat{\kappa}_2$ is the sample variance, as a unit of measurement in determining the placement of the u_q. As a rule of thumb, placing the two points closest to the origin at $\pm \frac{1}{4\sqrt{\hat{\kappa}_2}}$ and then gradually increasing the interval between consecutive points as $|u| \to \infty$ appears to work reasonably well. Figure 1 provides an indication of the agreement of the sample transform $\hat{y}_n(u)$ with the theoretical transform $y(u)$ for a sample of size 500. We thus expect to do reasonably well in estimating the parameters from the data with p=40.

In retrospect, it is believed that p could probably have been chosen to be somewhat less than 40 without seriously degrading the resultant estimators. If, however, the value p=40 is used in equation (3.9), it will be found that the sample covariance matrix $\hat{\underset{\sim}{K}}_n$ will be quite illconditioned. Its inversion may therefore sometimes prove to be numerically troublesome. A moderate (5 to 10 percent) inflation of its diagonal elements will alleviate this difficulty, and seems to have no harmful effect on the estimates.

There are theoretical difficulties associated with the fact that the function $0_n(\underset{\sim}{\theta})$ of equation (3.9) measures the deviation of the empirical function $\hat{y}_n(u)$ from the theoretical function $y(u,\underset{\sim}{\theta})$ at only a finite number of points. Although $y(u,\underset{\sim}{\theta})$ corresponds uniquely to the characteristic function $\phi(u,\underset{\sim}{\theta})$ and so uniquely determines the distribution function $F(x,\underset{\sim}{\theta})$ of equation (2.1) it may happen that two distinct feasible parameter vectors $\underset{\sim}{\theta}_1$ and $\underset{\sim}{\theta}_2$ satisfy $y(u_q,\underset{\sim}{\theta}_1) = y(u_q,\underset{\sim}{\theta}_2)$ at each of the points

$u_1, u_2, \ldots, u_p$ even though this cannot be the case identically in u. This is clearly undesirable since it implies that the estimation procedure is incapable of differentiating between samples drawn from the distinct distributions $F(x;\underset{\sim}{\theta}_1)$ and $F(x;\underset{\sim}{\theta}_2)$. Fortunately, from a practical point of view this phenomenon causes no difficulty as long as $u_1, u_2, \ldots, u_p$ are not too widely spaced, because then the vectors $\underset{\sim}{\theta}_1$ and $\underset{\sim}{\theta}_2$ must be greatly separated in the parameter space.

In order to make this last statement more precise, consider as a simple example the use of just six points located on the u-axis at $\pm \varepsilon$, $\pm 2\varepsilon$ and $\pm 3\varepsilon$, where $\varepsilon>0$. Then by straightforward algebra it can be shown that the six values $y(u,\underset{\sim}{\theta})$, $u = \pm \varepsilon, \pm 2\varepsilon, \pm 3\varepsilon$, uniquely determine the four elements of the parameter vector $\underset{\sim}{\theta} = (\mu,\lambda,\sigma_1^2,\sigma_2^2)'$ if $\underset{\sim}{\theta}$ is further assumed to lie in the reduced parameter space $\Theta_R = \{\underset{\sim}{\theta} \mid |\mu| < \frac{\pi}{\varepsilon}, \lambda|\mu| < \frac{\pi}{3\varepsilon}, \lambda,\sigma_1^2,\sigma_2^2>0\}$. To continue with this example, suppose the grid size ε is chosen equal to $\frac{1}{4\sqrt{\hat{\kappa}_2}}$, which is the position of the smallest positive point according to the previously mentioned rule of thumb, and suppose further that the sample variance $\hat{\kappa}_2$ estimates $\hat{\kappa}_2$ essentially without error. Then, letting $\underset{\sim}{\theta}_0 = (\mu_0,\lambda_0,\sigma_{10}^2,\sigma_{20}^2)$ denote the true values of the parameters,

$$\frac{|\mu_0|}{\sqrt{\hat{\kappa}_2}} \cong \frac{|\mu_0|}{\sqrt{\kappa_2}} = \frac{|\mu_0|}{\sqrt{\sigma_{10}^2+\lambda_0(\mu_0^2+\sigma_{20}^2)}} \leq \frac{1}{\sqrt{\lambda_0}} \tag{3.10}$$

and

$$\frac{|\mu_0|\lambda_0}{\sqrt{\hat{\kappa}_2}} \leq \sqrt{\lambda_0}. \tag{3.11}$$

Thus $\underset{\sim}{\theta}_0$ will lie in Θ_R as long as λ_0 is in the interval

$$\lambda_0 \in \left(\frac{1}{16\pi^2}, \frac{16\pi^2}{9}\right) \cong (0.006,\ 17.54) \tag{3.12}$$

since then, from equations (3.10) and (3.11)

$$|\mu_0| \leq \frac{\sqrt{\kappa_2}}{\sqrt{\lambda_0}} \leq \frac{\pi}{(1/4\sqrt{\kappa_2})} = \frac{\pi}{\varepsilon},$$

$$|\mu_0|\lambda_0 \leq \sqrt{\kappa_2}\sqrt{\lambda_0} \leq \frac{\pi}{(3/4\sqrt{\kappa_2})} = \frac{\pi}{3\varepsilon}.$$

For λ_0 in the interval (3.12), the fact that the values of $y(u,\underset{\sim}{\theta})$ at the six points $u=\pm\varepsilon, \pm 2\varepsilon, \pm 3\varepsilon$ completely specify $\underset{\sim}{\theta}$ within Θ_R allows one to construct a proof of the strong consistency of the estimates obtained by minimizing $O_n(\underset{\sim}{\theta})$ of equation (3.9) over any compact subset of Θ_R containing the true parameter vector. The proof is totally analogous to that given by Bryant and Paulson (1979). If, on the other hand, λ_0 does not lie in the interval (3.12) effective simultaneous estimation of all four parameters is a practical impossibility no matter what method may be used. The reasons for this phenomenon, which have to do with the insensitivity of the distribution function $F(x;\underset{\sim}{\theta})$ to its parameters for extreme values of λ, will be subsequently discussed.

The simple example of the preceding paragraphs is not meant to imply that only six u-values should be used in the computation of the objective function $O_n(\underset{\sim}{\theta})$, or even that $u_1, u_2, \ldots, u_p$ should be necessarily equally spaced. Rather, it is intended to at least partially justify the empirical observation that, even though the measurement of the deviation between the functions $\hat{y}_n(u)$ and $y(u;\theta)$ at only a finite number of points poses theoretical difficulties, these should not disqualify the proposed estimation procedure from practical consideration.

4. Maximum Likelihood Estimators

The probability density function of the modified compound Poisson distribution is given in equation (2.2) and may be expressed as

$$f(x) = \sum_{q=0}^{\infty} p_q f_q(x)$$

where

$$p_q = e^{-\lambda} \lambda^q / q!$$

and

$$f_q(x) = \frac{1}{(2\pi(\sigma_1^2+q\sigma_2^2))^{\frac{1}{2}}} \exp\left\{\frac{-(x-q\mu)^2}{2(\sigma_1^2+q\sigma_2^2)}\right\}.$$

In these equations the dependence of $f(x)$, p_q and $f_q(x)$ on the parameters has been suppressed for later convenience. A system of maximum likelihood equations can be formed by differentiating the log-likelihood function

$$L(\mu,\lambda,\sigma_1^2,\sigma_2^2) = \sum_{j=1}^{n} \log f(X_j)$$

with respect to the parameters μ, λ, σ_1^2, σ_2^2, and setting the resulting expressions equal to zero. This gives

$$\frac{\partial L}{\partial \mu} = \sum_j \frac{1}{f(X_j)} \sum_q \frac{q p_q}{\sigma_1^2+q\sigma_2^2} (X_j - q\mu) f_q(X_j) = 0 \qquad (4.1a)$$

$$\frac{\partial L}{\partial \lambda} = \frac{1}{\lambda} \sum_j \frac{1}{f(X_j)} \sum_q q p_q f_q(X_j) - n = 0 \qquad (4.1b)$$

$$\frac{\partial L}{\partial \sigma_1^2} = \tfrac{1}{2} \sum_j \frac{1}{f(X_j)} \sum_q \frac{p_q}{(\sigma_1^2+q\sigma_2^2)^2} \{(X_j-q\mu)^2 - (\sigma_1^2+q\sigma_2^2)\} f_q(X_j) = 0 \qquad (4.1c)$$

$$\frac{\partial L}{\partial \sigma_2^2} = \tfrac{1}{2} \sum_j \frac{1}{f(X_j)} \sum_q \frac{qp_q}{(\sigma_1^2+q\sigma_2^2)^2} \{(X_j-q\mu)^2 - (\sigma_1^2+q\sigma_2^2)\} f_q(X_j) = 0. \qquad (4.1d)$$

Explicit solution of these equations is, of course, impossible; however, it has been possible to express them in a form amenable to a fixed-point solution. It is convenient for this purpose to replace the parameter σ_1^2 by

$$\gamma = \frac{\sigma_1^2}{\sigma_2^2} .$$

Then if $\mu^{(m)}$, $\lambda^{(m)}$, $\sigma_2^{2(m)}$ and $\gamma^{(m)}$ represent the values of the estimates at the mth iteration of the fixed point algorithm, the equations (4.1) may be manipulated to suggest the following updating procedure:

$$\mu^{(m+1)} = \sum_j \frac{X_j}{f(X_j)} \sum_q \frac{qp_q f_q(X_j)}{(\gamma+q)} \Bigg/ \sum_j \frac{1}{f(X_j)} \sum_q \frac{q^2 p_q}{(\gamma+q)} f_q(X_j) \qquad (4.2a)$$

$$\lambda^{(m+1)} = \frac{1}{n} \sum_j \frac{1}{f(X_j)} \sum_q qp_q f_q(X_j) \qquad (4.2b)$$

$$\sigma_2^{2(m+1)} = \frac{1}{n} \{\sum_j \frac{X_j^2}{f(X_j)} \sum_q \frac{p_q}{(\gamma+q)} f_q(X_j) - \mu^2 \sum_j \frac{1}{f(X_j)} \sum_q \frac{q^2 p_q}{(\gamma+q)} f_q(X_j)\} \qquad (4.2c)$$

$$\gamma^{(m+1)} =$$

$$\frac{\gamma\{\sum_j \frac{X_j^2}{f(X_j)} \sum_q \frac{p_q f_q(X_j)}{q(\gamma+q)^2} - 2\mu \sum_j \frac{X_j}{f(X_j)} \sum_q \frac{qp_q}{q(\gamma+q)^2} f_q(X_j) + \mu^2 \sum_j \frac{1}{f(X_j)} \sum_q \frac{q^2 p_q}{q(\gamma+q)^2} f_q(X_j)\}}{\sigma_2^2 \sum_j \frac{1}{f(X_j)} \sum_q \frac{p_q f_q(X_j)}{(\gamma+q)}} \qquad (4.2d)$$

where the right hand side of (4.2) is evaluated at the mth values of the parameters.

There are an almost unlimited number of ways to transform the equations (4.1) into a form suitable for a fixed point procedure; the form of these equations is partially motivated by heuristic considerations. It has not been possible to prove that this fixed-point scheme must converge or yield unique solutions but the algorithm has consistently yielded reasonable parameter estimates when applied to simulated data.

The convergence of the iterates $\mu^{(m)}$, $\lambda^{(m)}$, $\sigma_2^{2(m)}$ and $\gamma^{(m)}$ of system (4.2) is unfortunately extremely slow, so that some sort of acceleration modification is a practical necessity. An adaption of the Aitken Δ^2 process (Hildebrand, 1974, pp. 567-71) has proven effective. A detailed description is given in Bryant (unpublished Ph.D. Thesis, Rensselaer Polytechnic Institute, 1977).

The major disadvantage of using the fixed-point algorithm to obtain maximum likelihood estimates for the parameters of the modified compound Poisson process is the inordinate amount of computer time required. For samples of size 500 the maximum likelihood procedure took

from two to six times as long as the characteristic function procedure, depending on the values of the parameters chosen (the amount of time required increased rapidly with λ). Furthermore, the time required by the maximum likelihood method increased with increasing sample size, so that for very large data sets its use may not be considered economically feasible. This is not the case with the characteristic function method, as the time it requires is primarily a function of p. As previously stated, the value of p used in the estimations reported here was 40, which was quite probably excessive; thus it might be possible to reduce the amount of computer time required by this procedure without loss in the accuracy of estimation.

5. Empirical Comparisons

Estimates of the parameters of simulated modified compound Poisson samples of size 500 are tabulated in Table 1, and may be used to at least partially evaluate the relative desirability of the cumulant matching, characteristic function and maximum likelihood estimation procedures. In view of the considerable amount of computer time required by the characteristic function and maximum likelihood algorithms, the number of different combinations of parameter values investigated was necessarily rather small; in all, the results of 25 simulated samples are contained in the table.

These data clearly indicate that cumulant matching estimates are noticeably less efficient than those provided by the other two procedures. In fact, in the majority of cases it was found that the cumulant matching

method gave infeasible solutions in which one or both of the estimates of the variance parameters σ_1^2 and σ_2^2 were negative. This is consistent with the results obtained by Press in his attempt to fit security price data with the modified compound Poisson distribution through use of this method of parameter estimation.

The characteristic function estimates and those obtained by maximum likelihood are highly correlated, and for most of the 25 samples in Table 1 yield solutions of nearly equal quality. Their comparison is made more difficult by the apparent fact that for some of the combinations of parameters considered (notably whenever $\mu=0$ and $\lambda\geq 1$), sample sizes considerably larger than 500 are required for truly effective estimation by any means. Accordingly, it would have been desirable to simulate larger samples with which to compare these procedures. This, however, was determined to be inadvisable due to the excessive amount of computer time it would have required. Instead, the parameters of some of the distributions were re-estimated using the same data as that upon which Table 1 is based, where in addition it was assumed that the ratio of the variance parameters σ_1^2 and σ_2^2 was known. This additional information increases the precision of estimation of both procedures, and also substantially reduces the length of time required by either algorithm. The resulting estimates are recorded in Table 2, along with those calculated from several samples of size 500 not included in the first data set. Conclusions similar to those drawn on the basis of Table 1 are supported by these data. Again, characteristic function and maximum likelihood estimators appear to do about equally well (an exception occurs in

TABLE 1

Comparison of Cumulant Matching (CM), Characteristic Function (CF) and Maximum Likelihood (MLE) Estimates of the Parameters of the Modified Compound Poisson Distribution

	μ	λ	σ_1^2	σ_2^2	μ	λ	σ_1^2	σ_2^2
Parameters	4	½	1	1	2	½	1	1
CM	5.145	0.403	0.248	-1.060	3.194	0.348	0.787	-0.408
CF	3.912	0.536	0.992	1.517	2.592	0.431	0.970	0.719
MLE	3.868	0.534	0.987	1.508	2.599	0.429	0.954	0.660
Parameters	1	½	1	1	0	½	1	1
CM	2.233	0.215	1.096	-0.159	-0.354	0.170	1.331	1.424
CF	1.395	0.338	1.210	0.841	-0.189	0.329	1.079	1.638
MLE	1.429	0.332	1.193	0.788	-0.123	0.535	1.061	1.020
Parameters	4	1	1	1	2	1	1	1
CM	3.488	1.121	1.574	3.154	3.233	0.682	0.637	-1.452
CF	3.956	0.993	1.079	1.435	1.362	1.620	0.617	2.157
MLE	4.011	0.975	1.192	1.318	1.396	1.578	0.669	2.097
Parameters	1	1	1	1	4	2	1	1
CM	1.257	0.683	1.009	0.720	4.135	1.943	-0.244	1.371
CF	1.039	0.825	0.947	0.906	4.122	1.946	1.302	0.470
MLE	1.179	0.727	0.988	0.804	4.127	1.944	1.254	0.513
Parameters	2	2	1	1	1	2	1	1
CM	0.773	5.210	-5.516	-2.506	2.292	0.901	0.557	-0.463
CF	2.375	1.696	0.620	0.135	0.877	2.118	1.068	0.704
MLE	2.371	1.701	0.578	0.153	1.178	1.755	1.165	0.527
Parameters	4	3	1	1	2	3	1	1
CM	6.063	1.381	-10.300	-5.022	3.07	1.94	-0.283	-1.04
CF	3.941	3.059	0.967	1.007	2.436	2.458	1.089	0.939
MLE	3.816	3.130	0.790	1.158	2.412	2.470	1.061	0.949
Parameters	1	3	1	1	-	½	½	1
CM	1.15	2.54	0.725	1.22	4.660	0.410	-0.006	-1.291
CF	1.145	2.549	1.294	0.941	4.068	0.478	0.238	0.778
MLE	1.207	2.418	1.339	0.896	4.064	0.472	0.242	0.814
Parameters	2	½	½	1	1	½	½	1
CM	2.626	0.332	0.246	-0.415	1.079	0.463	0.157	1.588
CF	2.404	0.371	0.271	0.353	1.116	0.449	0.249	1.304
MLE	2.403	0.370	0.274	0.347	1.147	0.438	0.253	1.280
Parameters	0	½	½	1	4	1	½	1
CM	0.051	0.355	1.426	-0.694	2.220	1.770	-1.560	5.400
CF	0.103	0.428	0.287	1.085	4.088	0.357	0.242	1.203
MLE	0.092	0.441	0.286	1.078	4.088	0.357	0.242	1.196

Table 1 continued

	μ	λ	σ_1^2	σ_2^2	μ	λ	σ_1^2	σ_2^2
Parameters	2	1	½	1	1	1	½	1
CM	2.977	0.601	-0.194	-1.211	1.509	0.744	0.368	0.967
CF	1.863	0.962	0.247	1.004	1.191	0.930	0.293	1.294
MLE	1.862	0.961	0.241	0.952	1.174	0.941	0.286	1.282
Parameters	0	1	½	1	4	2	½	1
CM	-0.001	78.705	12.872	-0.146	4.448	1.736	-0.456	0.617
CF	-0.165	0.439	0.480	1.699	3.996	1.939	0.195	0.871
MLE	-0.184	0.357	0.530	2.011	3.993	1.937	0.181	0.917
Parameters	2	2	½	1	1	2	½	1
CM	2.454	1.701	-0.358	0.020	2.012	1.012	0.360	-0.089
CF	1.573	2.497	0.037	1.154	0.969	2.108	0.342	0.966
MLE	1.834	2.278	0.112	0.833	0.931	2.189	0.318	0.969
Parameters	0	2	½	1				
CM	-0.114	0.493	1.310	2.285				
CF	-0.040	1.347	0.622	1.335				
MLE	-0.046	1.287	0.647	1.384				

the case where $\mu=4$, $\lambda=4$, $\sigma_1^2=\sigma_2^2 = 1$), and even when the variance parameters are assumed equal, effective estimation is not possible by either method when $\mu=0$, $\lambda=2$, $\sigma_1^2=\sigma_2^2 = 1$.

The inability to estimate accurately when $\mu=0$ and $\lambda\geq 1$ noted in both Tables 1 and 2 has been further substantiated empirically and may be explained by noting that the distribution function $F(x;\underset{\sim}{\theta})$ of equation (2.1) is nearly singular in this region when regarded as a function of its parameters. By this is meant that the distribution is strongly dependent only on certain combinations of the parameters and so is only very slightly perturbed as the parameter vector is allowed to vary on the hypersurfaces generated by fixing these combinations equal to some constants. In the case where $\mu=0$, the underlying distribution is symmetric, and its first three moments are matched by any symmetric member of the modified compound Poisson family whose parameters satisfy

$$\hat{\sigma}_1^2 + \lambda\hat{\sigma}_2^2 = \kappa_2 \tag{5.1}$$

where κ_2 is the population variance. Symmetric distributions whose parameters lie near the hypersurface (5.1) may therefore be virtually indistinguishable even though their parameters differ widely. For example, in the case of Table 1 where the true parameter values are $\mu=0$, $\lambda=2$, $\sigma_1^2=\frac{1}{4}$ and $\sigma_2^2=1$, each of the three estimation methods fits the data with a curve which is nearly symmetric and which has almost the same mean and variance as does the underlying population, as can be verified by the use of equations (2.4). These fitted distributions do not differ greatly from one another or from the parent distribution, and yet the corresponding

TABLE 2

Comparison of Characteristic Function (CF) and
Maximum Likelihood (MLE) Estimates of the Parameters of the
Modified Compound Poisson Distribution: σ_1^2 and σ_2^2 Assumed Equal

	μ	λ	σ_2^2	μ	λ	σ_2^2
Parameters	4	1	1	2	1	1
CF	3.980	0.986	1.112	2.176	1.013	1.126
MLE	4.018	0.973	1.201	2.119	1.035	1.165
Parameters	1	1	1	0	2	1
CF	1.004	0.854	0.931	0.054	0.383	2.340
MLE	1.030	0.833	0.931	0.021	0.416	2.298
Parameters	4	3	1	2	3	1
CF	3.946	3.055	0.983	1.916	3.281	1.046
MLE	3.842	3.123	0.922	1.867	3.189	1.182
Parameters	1	3	1	4	4	1
CF	1.000	2.918	1.055	3.936	4.786	0.919
MLE	1.034	2.824	1.055	3.945	3.824	0.823
Parameters	2	4	1	1	4	1
CF	2.423	3.301	0.513	1.084	3.390	1.034
MLE	2.425	3.311	0.609	1.110	3.806	1.028

estimates for the parameters λ, σ_1^2 and σ_2^2 are quite inaccurate.

The singularity of the modified compound Poisson distribution with $\mu=0$ is especially critical when λ is either large or else extremely small, for reasons which will be discussed shortly. If one is interested only in fitting a curve to data, then these considerations may not be of great concern. The precise estimation of the values of the individual parameters λ, σ_1^2 and σ_2^2 when μ is very close to zero, would, however, seem to require impractically large samples unless λ is of moderate magnitude.

Other interrelations among the parameters also exist and may cause difficulty in estimation. In some portions of the parameter space, for example, $F(x;\underset{\sim}{\theta})$ is quite insensitive to perturbations of the parameters on the surface

$$\lambda\mu = \kappa_1$$

where κ_1 is the population mean, and this in turn causes the estimators of λ and μ to be generally strongly negatively correlated, as is evident in both Tables 1 and 2. Other such relations are not so obvious. In an effort to gain some insight into these interrelations, and the effect they have on estimation, information matrices have been computed and are given in Bryant and Paulson (1981).

The information matrices reveal a very complicated pattern of interrelations among the parameters μ, λ, σ_1^2 and σ_2^2. No attempt will be made here to discuss these in complete detail, but rather only two observations, which will be of use in Section 6, will be noted. It may be seen that as λ increases, $F(x;\underset{\sim}{\theta})$ becomes nearly singular, which may be accounted for by the limiting normality of the standardized variate

$$\frac{X - \lambda\mu}{\sqrt{\sigma_1^2 + \lambda(\sigma^2+\sigma_2^2)}}$$

as $\lambda \to \infty$ while μ, σ_1^2 and σ_2^2 are kept fixed. The practical implication of this is that if λ is much larger than the values investigated in Table 1 only parametric combinations which are functions of $\lambda\mu$ and $\sigma_1^2 + \lambda(\mu^2+\sigma_2^2)$ may be accurately estimated without recourse to tremendously large samples. Conversely, if λ is very small the distribution is insensitive to the parameters μ and σ_2^2, since then the "noise" variate Z_0 of equation (1.1) is dominant. Finally, these matrices indicate that the magnitude of the variance σ_1^2 of Z_0 is a major factor in the overall estimability of the parameters. As seems reasonable, a small value of σ_1^2 relative to σ_2^2 permits more precise estimation than would be possible if this were not the case.

6. Designed Estimation Experiments

Suppose it is of interest to estimate the parameters $\underset{\sim}{\theta} = (\mu,\lambda,\sigma_1^2,\sigma_2^2)'$ of (1.2). If P(t) can be continuously and precisely observed, the arrival times of the Poisson process may be recorded along with the corresponding jumps of P(t), so that N(t), Y(t) and the Z_k are observable. In this case, estimation of the vector $\underset{\sim}{\theta}$ is straightforward. However, it is not difficult to think of applications where continuous observation is either not physically possible or else is not economically feasible, and yet the experimenter does possess some control over the times at which the process of equation (1.2) may be observed.

Suppose it has been decided that a fixed number n of observations of the process P(t) are to be used for purposes of estimation, and for simplicity it is agreed that a constant inter-observation time Δt will be employed. Then it is in the experimenter's interest to determine a value of Δt which will lead to parameter estimates of low variability. It should not be surprising to find that different choices will lead to estimates of varying quality, but the degree to which this is true turns out to be remarkable. If reasonable initial estimates of the true parameters are available, they may be used (in a manner to be described) to great advantage in the problem of selecting an appropriate inter-observation time.

Given some fixed Δt, the values $P(\Delta t)$, $p(2\Delta t),\ldots,P(n\Delta t)$ yield by differencing the random sample

$$X_j = P(j\Delta t) - P((j-1)\Delta t),\ j=1,2,\ldots,n,$$

where $P(0)=P_0$. These X_j have as their common distribution $F(x;\underset{\sim}{\theta}_{\Delta t})$, the modified compound Poisson distribution with the parameter vector

$$\underset{\sim}{\theta}_{\Delta t} = (\mu,\ \lambda_{\Delta t},\ \sigma^2_{1\Delta t},\ \sigma^2_2),$$

where $\lambda_{\Delta t} = \lambda\Delta t$ and $\sigma^2_{1\Delta t} = \sigma^2_1\Delta t$. The problem of estimating the parameters $\underset{\sim}{\theta}$ of the stochastic process is then reduced to that of estimating $\underset{\sim}{\theta}_{\Delta t}$; if any of the procedures discussed in Sections 2-4 is used to obtain the estimates

$$\hat{\underset{\sim}{\theta}}_{\Delta t} = (\hat{\mu},\ \hat{\lambda}_{\Delta t},\ \hat{\sigma}^2_{1\Delta t},\ \hat{\sigma}^2_2),$$

then the vector of process parameters is likewise estimated by

$$\hat{\underset{\sim}{\theta}} = (\hat{\mu},\ \hat{\lambda},\ \hat{\sigma}^2_1,\ \hat{\sigma}^2_2),$$

where $\hat{\lambda} = \hat{\lambda}_{\Delta t}/\Delta t$ and $\hat{\sigma}_1^2 = \sigma_{1\Delta t}^2/\Delta t$.

It was seen in Section 5 that the character of $F(x;\underset{\sim}{\theta}_{\Delta t})$ with regard to the estimability of its parameters is strongly influenced by the magnitudes of $\lambda_{\Delta t}$ and $\sigma_{1\Delta t}^2$. These quantities may be controlled by the experimenter with proper selection of the inter-observation time Δt, which in turn will permit accurate estimation. In fact, if this is not done, the near singularity of the information matrices over much of the parameter space makes it clear that accurate estimation may not be possible.

Let $\tilde{\underset{\sim}{\theta}}$ be a vector of prior estimates of the process parameters, and let $\tilde{\underset{\sim}{\theta}}_{\Delta t}$ be the corresponding initial estimates of $\underset{\sim}{\theta}_{\Delta t}$. If it is assumed momentarily that the estimates $\tilde{\underset{\sim}{\theta}}_{\Delta t}$ have a covariance matrix not too dissimilar to $\frac{1}{n} I(\tilde{\underset{\sim}{\theta}}_{\Delta t})^{-1}$ where the jk element of $I(\underset{\sim}{\theta})$ is given by

$$I_{jk}(\underset{\sim}{\theta}_{\Delta t}) = E\left\{ \frac{\partial \log f(x;\underset{\sim}{\theta}_{\Delta t})}{\partial \theta_{\Delta tj}} \; \frac{\partial \log f(x;\underset{\sim}{\theta}_{\Delta t})}{\partial \theta_{\Delta tk}} \right\}, \qquad (6.1)$$

then these information matrices may be used in conjunction with the initial estimates to approximate, for any given Δt, the quantities $nVar(\hat{\mu})$, $nVar(\hat{\lambda}) = nVar(\hat{\lambda}_{\Delta t})/\Delta t^2$, $nVar(\hat{\sigma}_1^2) = nVar(\hat{\sigma}_{1\Delta t}^2)/\Delta t^2$ and $nVar(\hat{\sigma}_2^2)$. An approximation of the generalized variance of $\hat{\underset{\sim}{\theta}}$ may also be useful, and is provided by the quantity $\{|I(\tilde{\underset{\sim}{\theta}}_{\Delta t})|\Delta t^4\}^{-1}$. By calculating these numbers for several values of Δt, it is possible to select an interobservation time which permits near-optimal performance in terms of the overall variability of the estimates of the process parameters.

It should be mentioned here that $\frac{1}{n} I(\tilde{\underset{\sim}{\theta}}_{\Delta t})^{-1}$ may not be a very good approximation for $Var(\hat{\underset{\sim}{\theta}}_{\Delta t})$ even if maximum likelihood is used to generate these estimators and the number of observations is large, since the information matrices are typically ill-conditioned so that a very slow approach to the asymptotic distribution must be expected. This, however,

does not invalidate the procedure since the information matrices may still be regarded as indicators of estimability, at least as long as the prior vector estimate $\tilde{\theta}$ is not too inaccurate.

To illustrate the suggested method, suppose that a process P(t) has true parameter values $\mu=1$, $\lambda=5$/unit time, $\sigma_1^2=\frac{1}{2}$/unit time and $\sigma_2^2=1$ which are to be estimated based on 500 observations. If the computations indicated above are performed for several values of Δt (using the true parameters as initial estimates) the data of Table 3 will result. There is a clear indication that an inter-observation time of roughly 0.1 of one time unit will produce nearly optimal estimates of the process parameters. On the other hand, if Δt were to be arbitrarily chosen to be 1.0 (resulting in an experiment which would run ten times as long, with sampling only one-tenth as often) effective estimation would be impossible without dramatically increasing the number of observations.

Four simulated samples of size 500, whose parameters correspond to inter-observation times of 1.0, 0.2, 0.1 and 0.05 time units, were generated and their parameters estimated by means of the characteristic function algorithm. The same stream of random numbers was used in the creation of each of the samples to facilitate the comparison of data insofar as possible. Results are displayed in Table 4, and these show that reasonable estimates are achieved for Δt in the range from 0.05 to 0.2, while as predicted the solution obtained for an inter-observation time of 1.0 is quite inaccurate.

TABLE 3

Effect of Inter-Observation Time
on Parameter Estimation
$\mu=1$ $\lambda=5$ $\sigma_2^2=0.5$ $\sigma_2^2=1$

Δt	μ	$\lambda_{\Delta t}$	$\sigma^2_{1\Delta t}$	σ_2^2	nVar($\hat{\mu}$)	nVar($\hat{\lambda}_{\Delta t}$)	nVar($\hat{\lambda}$)	nVar($\hat{\sigma}^2_{1\Delta t}$)	nVar($\hat{\sigma}_1^2$)	nVar($\hat{\sigma}_2^2$)
1.0	1	5	0.5	1	112.92	2824.7	2824.7	1446.3	1446.3	69.591
0.2	1	1	0.1	1	6.0847	6.1373	154.33	0.20932	5.2330	7.7650
0.1	1	0.5	0.05	1	5.3150	1.3424	134.24	0.018248	1.8248	9.1207
0.05	1	0.25	0.025	1	5.9360	0.43382	173.53	0.0024967	0.99869	12.959

Δt	μ	$\lambda_{\Delta t}$	$\sigma^2_{1\Delta t}$	σ_2^2	$\lvert I(\underline{\theta}_{\Delta t})\rvert^{-1}$	$\lvert I(\underline{\theta}_{\Delta t})\rvert^{-1}/(\Delta t)^4$
1.0	1	5	0.5	1	0.278×10^{7}	0.278×10^{7}
0.2	1	1	0.1	1	0.115×10^{2}	0.716×10^{4}
0.1	1	0.5	0.05	1	0.573×10^{0}	0.573×10^{4}
0.05	1	0.25	0.025	1	0.724×10^{-1}	0.116×10^{5}

TABLE 4

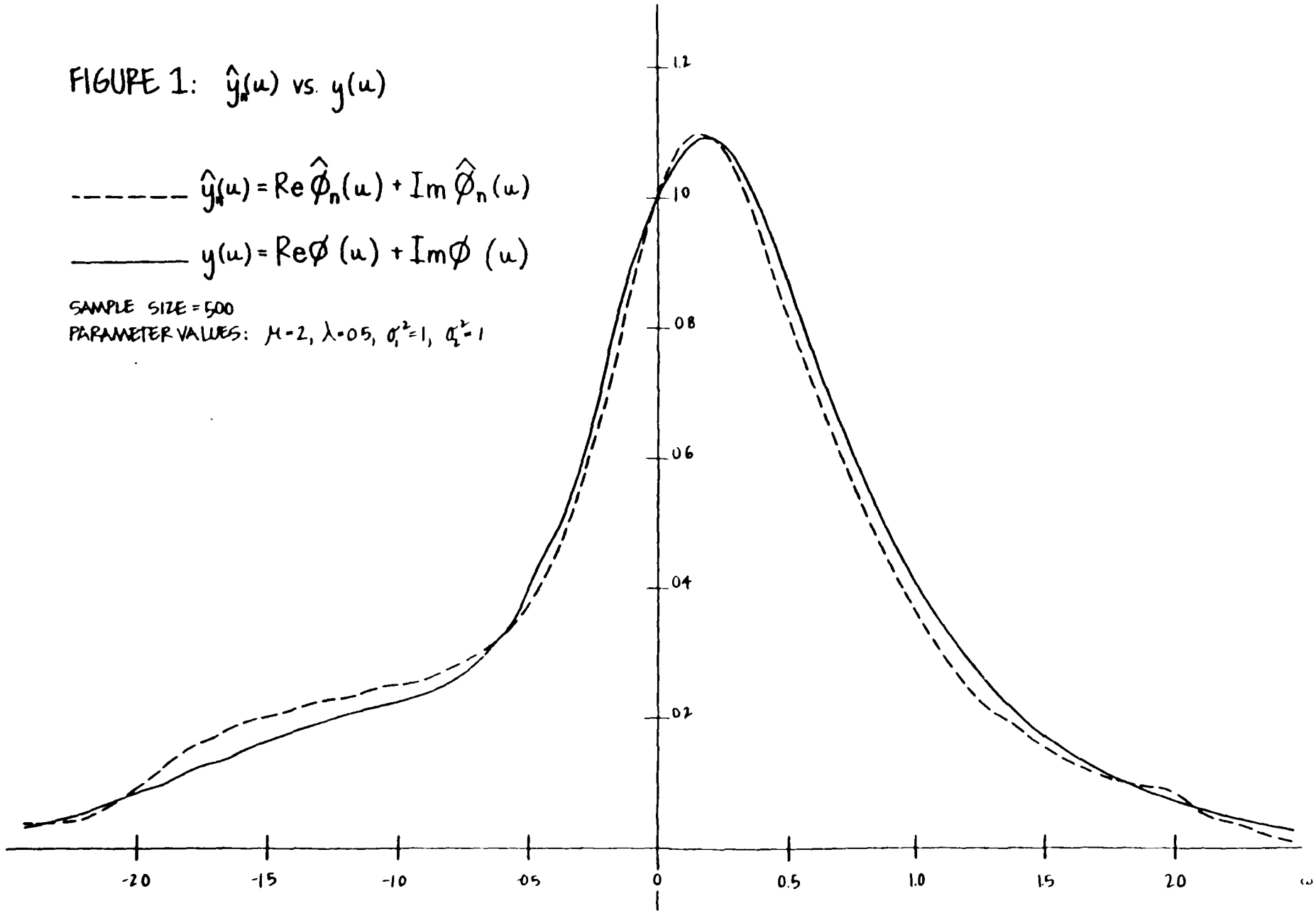
Effect of Inter-Observation Time on
Parameter Estimation - Simulation Results

μ=1 λ=5 $\hat{\sigma}_1^2$=0.5 σ_2^2=1 n=500

Δt	$\hat{\mu}$	$\hat{\lambda}_{\Delta t}$	$\hat{\sigma}^2_{1\Delta t}$	$\hat{\sigma}_2^2$	$\hat{\mu}$	$\hat{\lambda}$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$
1.0	1.646	3.042	1.781	0.000	1.646	3.042	1.781	0.000
0.2	0.900	1.0786	0.0895	1.065	0.900	5.393	0.448	1.065
0.1	0.938	0.5282	0.0465	1.047	0.938	5.282	0.465	1.047
0.05	1.014	0.2241	0.0252	0.933	1.014	4.482	0.504	0.933

While it is true that Δt was selected with prior knowledge of the true parameters in this example, it is also evident that it need not be chosen with extreme precision in order to gain acceptable results. Thus it is believed that this method can be used with success in conjunction with reasonable initial estimates. A simple, useful rule of thumb is the following: if the arrival rate is λ per unit time, then the sampling rate should be $(2\lambda)^{-1}$ per unit time. The use of such a selection of inter-observation time can lead to much enhanced parameter estimates.

FIGURE 1: $\hat{y}_n(u)$ vs. $y(u)$



References

Bryant, J.L. and Paulson, A.S. (1981). Estimation of the parameters of mixtures via distance between densities or characteristic functions. Operations Research and Statistics Report 37-81-P9, Rensselaer Polytechnic Institute.

Bryant, J.L. and Paulson, A.S. (1979). Some Comments on Characteristic Function-Based Estimators, Sankhyā, Series A, 41, 109-116.

Fama, E.F., "The Behavior of Stock Market Prices," Journal of Business, 38 (1965), pp. 34-105.

Feller, W. (1968). An Introduction to Probability and Its Applications, Vol. II, New York: Wiley.

Feuerverger, A. and McDunnough, P. (1979). On Some Fourier Methods for Inference. To appear in the Journal of the American Statistical Association.

Feuerverger, A. and McDunnough, P. (1980). On the Efficienty of Empirical Characteristic Function Procedures, Journal of the Royal Statistical Society, Series B, 43, pp. 20-27.

Fielitz, B.D. and Smith, E.W., "Asymmetric Stable Distributions of Stock Price Changes," J. Amer. Statist. Assoc., 67 (1972), pp. 813-814.

Hildebrand, F.B., Introduction to Numerical Analysis. New York: McGraw-Hill, 1974.

Jacoby, S.L.S., Kowalik, J.S. and Pizzo, J.T., Iterative Methods for Nonlinear Optimization Problems. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Kendall, M.G. and Stuart, A., The Advanced Theory of Statistics, Vol I, New York: Hafner, 1969.

Leitch, R.A. and Paulson, A.S., "Estimation of Stable Law Parameters: Stock Price Behavior Application," J. Amer. Statist. Assoc., 70 (1975), pp. 690-697.

Press, S.J., "A Compound Events Model for Security Prices," Journal of Business, 40 (1967), p. 217-235.

Press, S.J., "A Modified Compound Poisson Process with Normal Compounding," J. Amer. Statist. Assoc., 63 (1968), pp. 607-613.

Thornton, J.C. and Paulson, A.S., "Asymptotic Distribution of Estimators for the Stable Laws," to appear in Sankhyā, Series A, (1977).

MASTER COPY - FOR REPRODUCTION PURPOSES

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO-2	2. GOVT ACCESSION NO. N/A AD-A099510	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) Estimation of the Parameters of a Modified Compound Poisson Distribution		5. TYPE OF REPORT & PERIOD COVERED Working Paper 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) J.L. Bryant A.S. Paulson		8. CONTRACT OR GRANT NUMBER(s) DAAG29-81-K-0110
9. PERFORMING ORGANIZATION NAME AND ADDRESS Rensselaer Polytechnic Institute Troy, New York 12180		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE 13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Characteristic function, empirical characteristic function, compound Poisson, maximum likelihood, normal variables

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper addresses the problem of estimation of the parameters of the Poisson sum of Gaussian random variables imbedded in a background of Gaussian noise when only realizations of the sum are observable. Cumulant matching, maximum likelihood, and an empirically orthogonalized characteristic function procedures are considered. The characteristic function and the maximum likelihood procedures produce similar results in a simulation study. However, the characteristic function procedure is computationally superior. Conditions under which all procedures are incapable of parameter estimation are discussed.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED


DATE
FILMED
5-8